RAGNAR FRISCH

ON THE INVERSION OF A MOVING AVERAGE

Reprinted from the Skandinavisk Aktuarietidskrift 1938

UPPSALA 1938
ALMQVIST & WIKSELLS BOKTRYCKERI-A.-B.

On the Inversion of a Moving Average.

By **Ragnar Frisch** (Oslo).

At the request of the Faculty of Science of Stockholms Högskola I had the pleasure to act as opponent at the doctoral discussion on Mr. HERMAN WOLD's dissertation: »A Study in the Analysis of Stationary Time Series». Both the reading of the book, the public discussion and a subsequent private correspondance with the author, I have found very interesting and stimulating. I have also had the pleasure to discuss to some extent this matter with Professor CRAMÉR. Some of the points raised during our exchanging of views have been covered by Mr. WOLD in his note in this issue of the »Aktuarietidskrift», but not all. It may therefore be worth while to add a few remarks. I need not dwell upon the various merits of the book, they speak for themselves. Here, I shall confine myself to one particular point where some difference of opinion still seems to persist, namely the significance of the formula (255) in the dissertation, which is the same as (13) in Mr. WOLD's note in this issue. This formula is intended as an inversion formula for a moving average on a random variable in the singular case where the characteristic equation of the moving average has at least one root on the unit circle.

The right member of the formula in question is a double limiting process, I shall therefore prefer to write it explicitely

$$(1) \qquad \eta_t = \lim_{n \to \infty} \lim_{N \to \infty} \left(a_0^{(n)} \varsigma_t + a_1^{(n)} \varsigma_{t-1} + \cdots + a_N^{(n)} \varsigma_{t-N} \right)$$

I omit the brackets {} since it is unessential for the present argument whether we consider a given time series, or a random process that is capable of producing time series of a given sort. The limiting process in N — which is to be carried out before the limiting process in n — does not converge uniformely in n, hence the order of the two limiting processes in the formula cannot be interchanged.

The formula as it stands does therefore *not* tell us what sort of numerical computations to carry out in an actual case. In view of the non-uniformity of the convergency the approximation obtained by interrupting the double limiting process will depend fundamentally on the particular *way* in which the number of steps in one direction is fixed as compared with the number of steps in the other direction. Mr. WOLD has not given any discussion of how this fact works out in the present case, not even a rough estimate of the remainder is given. In my opinion it is therefore not justified to consider the numerical computation which he gives in (263) of the book as an example of the »application» of (253).

In practice the problem is rather the reverse of that indicated by (1). In an actual case it will be more relevant to start with a given finite N. But then the n-limiting process becomes *entirely unnecessary*. One would then only ask for some particular a-set to be associated with the given N (possibly a set independent of N). The double limiting process and all the difficulties that go with it are thus avoided. The fact that such a simple analysis is possible and even yields considerably *more* than what could be obtained by Mr. WOLD's method, even if it were carefully worked out with remainder term, indicates, it seems to me, that the auxiliary b-sequence of Mr. WOLD is a very *artificial* procedure in the present problem.

The following is a discussion of the problem along what seems to me to be more fruitful lines[1]. First consider as an example the simple case

(2) $$\varsigma_t = \eta_t - \eta_{t-1}.$$

[1] This was worked out on my way back from Stockholm Oct. 5. 1938, except the final form of the solution of the minimalization problem, which was shaped subsequently.

Let N be a given number and suppose that $\varsigma_t, \varsigma_{t-1}, \ldots \varsigma_{t-N}$ are known. On the basis of this we want to estimate η_t. Let k be some real constant and consider

$$(3) \qquad \eta_t^* = \varsigma_t + k\,\varsigma_{t-1} + \cdots + k^N \varsigma_{t-N}.$$

To select the coefficients of the linear ς-form exponentially as in (3) is by no means the most effective procedure, but we start by this so show exactly what is involved in Mr. WOLD's numerical example (263).

Consider the difference

$$(4) \qquad \delta_t = \eta_t^* - \eta_t.$$

It is identically equal to

$$(5) \qquad \delta_t = -(1-k)(\eta_{t-1} + k\eta_{t-2} + \cdots + k^{N-1}\eta_{t-N}) - k^N \eta_{t-(N+1)}.$$

Therefore, if all the $\eta - s$ are auto-noncorrelated random variables with the same mean and variance, we get

$$(6) \qquad E\delta_t = -E\eta$$

$$(7) \qquad E(\delta_t - E\delta_t)^2 = \frac{1-k) + 2k^{2N+1}}{1+k} E(\eta - E\eta)^2.$$

From (7) is seen what function of N k would have to be in order to make the dispersion of δ_t the smallest. Further (7) shows that for any function $k(N)$ such that simultaneously $k \to 1$ and $k^N \to 0$ when $N \to \infty$, the dispersion in question tends towards zero, (if the variance of η is finite) and hence η_t^* becomes a converging estimate of $\eta_t - E\eta$ (when $E\eta$ is known). In a recent letter Professor Cramér suggested $k = 1 - \frac{1}{\sqrt{N}}$. It follows immediately from the above that this gives a converging estimate, indeed

$$k^N = \left[\left(1 - \frac{1}{\sqrt{N}}\right)^{\sqrt{N}}\right]^{\sqrt{N}} \to e^{-\sqrt{N}} \to 0 \qquad \text{as } N \to \infty.$$

No information about the nature of the function $k(N)$ is contained in Mr. WOLD's analysis, but without this information a computation for a finite N (as done in his example (263)) will not have a precise meaning.

Even with the complete information given by (7), the exponential selection of the coefficients is not very *effective* in the sense of yielding a small dispersion of the estimate when N is given. The most effective selection in the general case of an n-th order moving average, is determined as follows.

Let

$$(8) \qquad \varsigma_t = b_0 \eta_t + b_1 \eta_{t-1} + \cdots + b_h \eta_{t-h}$$

Let N be a given number and consider with some set of coefficients, $a_0, a_1, \ldots a_N$ the linear form

$$(9) \qquad \eta_t^* = a_0 \varsigma_t + a_1 \varsigma_{t-1} + \cdots + a_N \varsigma_{t-N}.$$

The difference

$$(10) \qquad \delta_t = \eta_t^* - a_0 b_0 \eta_t$$

is identically equal to

$$(11) \qquad \delta_t = c_1 \eta_{t-1} + c_2 \eta_{t-2} + \cdots + c_{N+h} \eta_{t-(N+h)}$$

where

$$(12) \qquad \begin{aligned} c_1 &= b_0 a_1 + b_1 a_0 \\ c_2 &= b_0 a_2 + b_1 a_1 + b_2 a_0 \\ &\cdots\cdots\cdots\cdots \\ c_{h-1} &= b_0 a_{h-1} + b_1 a_{h-2} + \cdots + b_{h-1} a_0 \end{aligned}$$

$$(13) \qquad c_x = b_0 a_x + b_1 a_{x-1} + \cdots + b_h a_{x-h}, \quad \text{for } x = h, h+1, \ldots N$$
$$(h \text{ assumed} \gtreqless N)$$

$$(14) \qquad \begin{aligned} c_{N+1} &= b_1 a_N + b_2 a_{N-1} + \cdots + b_h a_{N-h+1} \\ c_{N+2} &= b_2 a_N + b_3 a_{N-1} + \cdots + b_h a_{N-h+2} \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ c_{N+h} &= b_h a_N \end{aligned}$$

If the η's are auto-noncorrelated random variables with the same mean and variance we get

$$(15) \qquad E\delta_t = (c_1 + c_2 + \cdots + c_{N+h})\,E\eta = \\ = [(b_0 + b_1 + \cdots + b_h)(a_0 + a_1 + \cdots + a_N) - a_0 b_0]\,E\eta$$

$$(16) \qquad E(\delta_t - E\delta_t)^2 = (c_1^2 + c_2^2 + \cdots + c_{N+h}^2)\,E(\eta - E\eta)^2.$$

The problem is to minimize the parenthesis to the right in (16) where the individual terms are given by (12) — (14). These terms may all be looked upon as of the form (13) provided we introduce the side conditions

$$(17) \qquad a_{-1} = a_{-2} = \cdots = a_{-(h-1)} = 0$$

$$(18) \qquad a_{N+1} = a_{N+2} = \cdots = a_{N+h} = 0$$

(For $h = 1$ the set of conditions (17) disappears). In other words the problem is, with the side conditions (17) and (18), to minimize

$$(19) \qquad \Phi = \sum_{z=1}^{N+h} (b_0 a_z + b_1 a_{z-1} + \cdots + b_h a_{z-h})^2 = \\ = a_0^2(\mu_0 - b_0^2) + 2a_0 \sum_{x=1}^{N} a_x \mu_x + \sum_{x=1}^{N} \sum_{y=1}^{N} a_x a_y \mu_{x-y}$$

where

$$\mu_\tau = \mu_{-\tau} = \sum_{z=-\infty}^{+\infty} b_z b_{z+\tau} = b_0 b_{|\tau|} + b_1 b_{|\tau|+1} + \cdots + \\ + b_{h-|\tau|} b_h = \text{automoment of } b$$

$$(\mu_\tau = 0 \text{ for } |\tau| > h).$$

In addition we need some side condition which will exclude the trivial solution $a_0 = a_1 = \cdots = a_N = 0$. The selection of such a condition is essentially of the same nature as the selection of a given *direction* in which to take least squares in the regression theory of $(h + 1)$ variables. We adopt the condition

$$(20) \qquad a_0 = 1$$

Thus $a_1, a_2 \ldots a_N$ are the independent coefficients to dispose of. From (19) we get

$$(21)\qquad \frac{1}{2}\frac{\partial \Phi}{\partial a_x} = \sum_{\tau=-\infty}^{+\infty} \mu_\tau a_{x+\tau} = \mu_0 a_x + \mu_1 (a_{x+1} + a_{x-1}) + \cdots + \mu_h (a_{x+h} + a_{x-h})$$

$$(x = 1, 2 \ldots N)$$

This shows that *the extremal series a_x must satisfy the $2h-th$ order difference equation*

$$(22)\qquad \mu_0 a_x + \mu_1 (a_{x+1} + a_{x-1}) + \cdots + \mu_h (a_{x+h} + a_{x-h}) = 0$$

with the $2h$ initial conditions (17), (18) *and* (20). This determines the a_x series completely.

That this series actually furnishes a minimum is seen by noticing that

$$(23)\qquad \frac{1}{2}\frac{\partial^2 \Phi}{\partial a_x \partial a_y} = \mu_{y-x}.$$

Apart from the factor $\frac{1}{2}$ the matrix of the second order derivatives for $x = 1, 2 \ldots N$, $y = 1, 2 \ldots N$ (the independent a-coefficient) is thus a moment matrix namely the moment matrix over z from $-\infty$ to $+\infty$ of the N real functions of z, $b_{z+1}, \ldots b_{z+N}$, hence the matrix considered is positive definite.

The extremal conditions may also be written

$$\sum_{z=1}^{N} a_z \mu_{x-z} = - a_0 \mu_x \quad (x = 1, 2, \ldots N, \text{ but not } x = 0)$$

Inserting this in (19) we see that the extremal value of Φ is

$$(24)\qquad \Phi_{\min} = a_0 \sum_{x=0}^{N} a_x \mu_x - a_0^2 b_0^2.$$

The above solution is applicable to the inversion of a finite order moving average of a random variable no matter how the characteristic roots of the weight system of the average are distributed, inside, outside or on the unit circle.

Example I.

$$\varsigma_t = \eta_t - \eta_{t-1}$$

Here $h = 1$, $\mu_0 = 2$, $\mu_1 = -1$. The difference equation is $a_{x+1} - 2a_x + a_{x-1} = 0$ i. e. $\triangle^2 a_x = 0$. Apart from an arbitrary function of period 1, which is here unessential, a_x must consequently be linear in x. Further $a_0 = 1$, $a_{N+1} = 0$, that is

$$a_x = 1 - \frac{x}{N+1}. \tag{25}$$

Finally $E\delta_t = [(1-1)(a_0 + \cdots + a_N) - 1]\,E\eta = -E\eta$. Therefore

$$\eta_t^* = \varsigma_t + \left(1 - \frac{1}{N+1}\right)\varsigma_{t-1} + \cdots + \frac{1}{N+1}\varsigma_{t-N} \tag{26}$$

is the most effective estimate of $\eta_t - E\eta$ when N is given. Since $\Phi_{\min} = 2 - \left(1 - \frac{1}{N+1}\right) - 1 = \frac{1}{N+1}$, the estimate converges as $N \to \infty$.

Example II.

$$\varsigma_t = \eta_t - 2\eta_{t-1} + \eta_{t-2}.$$

Here $h = 2$, $\mu_0 = 6$, $\mu_1 = -4$, $\mu_2 = 1$. The difference equation is $a_{x+2} - 4a_{x+1} + 6a_x - 4a_{x-1} + a_{x-2} = 0$, i. e. $\triangle^4 a_x = 0$. Consequently (apart from the unessential periodic function) a_x is a third degree polynomial satisfying $a_{-1}, = a_{N+1} = a_{N+2} = 0$, $a_0 = 1$, that is

$$a_x = (x+1)\left(1 - \frac{x}{N+1}\right)\left(1 - \frac{x}{N+2}\right). \tag{27}$$

Finally $E\delta_t = ((1-2+1)(a_0 + \cdots + a_N) - 1)\,E\eta = -E\eta$. Therefore

$$\eta_t^* = \varsigma_t + 2\left(1 - \frac{1}{N+1}\right)\left(1 - \frac{1}{N+2}\right)\varsigma_{t-1} + \cdots + \frac{2}{N+2}\varsigma_{t-N} \tag{28}$$

is now the most effective estimate of $\eta_t - E\eta$. Since $\Phi_{\min} = 6 - 4.2\left(1 - \frac{1}{N+1}\right)\left(1 - \frac{1}{N+2}\right) + 3\left(1 - \frac{2}{N+1}\right)\left(1 - \frac{2}{N+2}\right) - 1 = \frac{2}{N+1} + \frac{2}{N+2} + \frac{4}{(N+1)(N+2)}$, the estimate converges as $N \to \infty$.

In both examples the most effective estimate is the same as that obtained by solving the difference equation in question by *an elementary* — and indeed exceedingly simple — *recurrence process.* For instance in the example I the classical recurrence scheme gives

$$\eta_T = S_T + C \tag{29}$$

where T is any of the points $t, t-1, \ldots, t-(N+1)$, C a constant independent of T, and S_T the known function

$$S_T = \varsigma_T + \varsigma_{T-1} + \cdots + \varsigma_{t-N} \quad \text{(conventionally } S_{t-(N+1)} = 0) \tag{30}$$

Thus, if the ς-series is known, the η-series is also known, (not only in the stochastical sense but exactly) apart from an additive constant. The problem of estimating any *particular η is therefore simply a question of estimating the constant in question.* This estimate follows immediately from (29) if as an estimate of $E\eta$ we take the average of η_T over all the $(N+1)$ available T values before t. This gives

$$C \text{ probably} = E\eta - \frac{S_{t-1} + \cdots + S_{t-(N+1)}}{N+1} =$$

$$= E\eta - \frac{\varsigma_{t-1} + 2\varsigma_{t-2} + \cdots + N\varsigma_{t-N}}{N+1}$$

and hence

$$\eta_t - E\eta \text{ probably} = \varsigma_t + \left(1 - \frac{1}{N+1}\right)\varsigma_{t-1} + \cdots + \left(1 - \frac{N}{N+1}\right)\varsigma_{t-N} \tag{31}$$

which is the same estimate of $\eta_t - E\eta$ as (26).

It should be noticed that it is not possible to arrive at a set up of the form here used, with a finite N, (which afterwards may be made the object of simple limiting process) by starting as Mr. Wold does from an auxiliary sequence $b_0^{(n)}, b_1^{(n)} \ldots b_h^{(n)}$ and inserting this in his (255).
